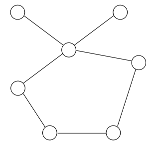
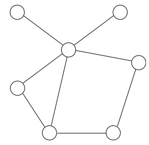
Probabilistic Graphical Models
Christoph Lampert <chl@ist.ac.at>
TA: all of us <pgm_2016@lists.ist.ac.at>
Exercise Sheet 5 v1.0

1 The treewidth of a graph

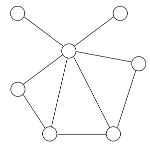
A crucial property to identify if probabilitic inference in a graphical model can be done efficiently is the *treewidth* of the underlying graph.



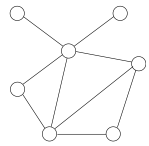
a) graph that is not chordal



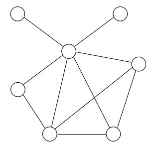
b) not a chordal completion of a)



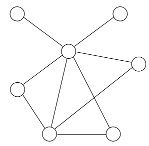
c) a chordal completion of a)



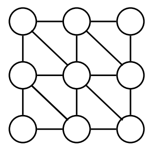
d) a chordal completion of a)



e) a chordal completion of a)



f) chordal, but not a completion of a)



h) not chordal (try to see why!)
If you really can't find the solution, there's a hint at the bottom of the page.

Definition 1. (see Wikipedia: `https://en.wikipedia.org/wiki/Chordal_graph`)
A graph is called *chordal*, if any cycle in it that consists of four or more vertices has a *chord*, i.e. there exists an edge that is not part of the cycle but connects two vertices of the cycle.

Definition 2. (see Wikipedia: `https://en.wikipedia.org/wiki/Chordal_completion`)
A *chordal completion* of a graph is a chordal graph that has the same vertex set and contains at least all edges of the original graph. Note: in general, graphs can have many different chordal completions.

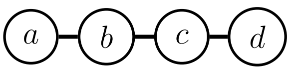
Definition 3. (see Wikipedia: `https://en.wikipedia.org/wiki/Treewidth`)
The *treewidth* of a chordal graph is the size of its largest clique minus 1. The *treewidth* of a (potentially non-chordal) graph is the smallest treewidth of any of its chordal completions.

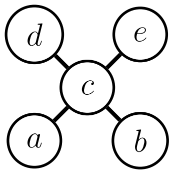
For each the following graphs 1)–7),

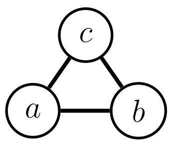
a) determine if it is chordal,

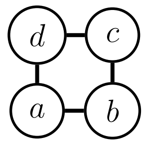
b) if not, construct a chordal completion,

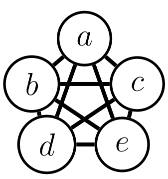
c) determine its treewidth.

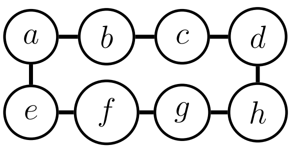
1) 

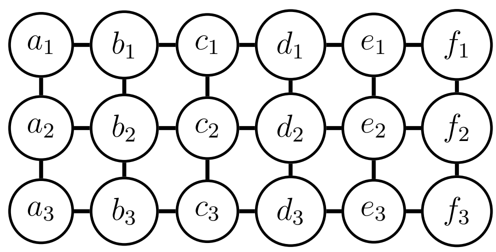
2) 

3) 

4) 

5) 

6) 

7) 

Hint for 1h): The big outer cycle has no chord.

2 Factor Graphs

Assume you are given eight binary-valued random variables, $X_1, \dots, X_8$. Construct factor graphs for the following probability distributions (with $x = (x_1, \dots, x_8)$), such that their underlying graphs have minimal treewidth.

a) $p(x) \propto e^{\text{number of 1s in } x}$

b) $p(x) \propto e^{\text{number of (0–1) transitions in } x_1,\dots,x_8}$

c) $p(x) \propto e^{\text{number of (0–1–0) transitions in } x_1,\dots,x_8}$

d) $p(x) \propto e^{\text{number of (1–0–1) combinations between any three distinct entries in } x}$

e) $p(x) = \begin{cases} 1 & \text{if } x = (0, 0, \dots, 0) \\ 0 & \text{otherwise} \end{cases}$

f) $p(x) = \begin{cases} \frac{1}{2} & \text{if } x = (1, 1, \dots, 1) \\ \frac{1}{510} & \text{otherwise} \end{cases}$

g) $p(x) \propto e^{\text{parity of } x}$

Could you do better, if you introduced additional (latent) random variables?

3 Marginal Inference

Assume you are given four binary-valued random variables, $X_1, \dots, X_4$, and a distribution $p(x_1, x_2, x_3, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_3, x_4)$ with factors $\phi_i(x_i, x_{i+1}) = \begin{cases} 3 & \text{if } x_i = 0 \text{ and } x_{i+1} = 1 \\ 1 & \text{otherwise} \end{cases}$ for $i = 1, \dots, 3$.

Compute (on paper!):

a) the normalizing constant

b) the probability $p(x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0)$

c) the marginal probability $p(x_1 = 0)$

d) $\text{corr}(X_1, X_2)$

e) the marginal probability $p(x_1 = 0, x_4 = 0)$

In each case, perform the computation in two ways: once naively, and once using belief propagation where possible (note: e) might require some thought for this). What is more efficient? How would this change *f*) for a larger number of variables, *g*) for variables with more states?

4 Maximum Entropy Distribution

Complete the proof we skipped in the lecture:

For samples $z^1, \dots, z^N$ and feature functions $\phi_i : \mathcal{Z} \to \mathbb{R}$ for $i = 1, \dots, d$, define $\mu_i := \sum_{n=1}^{N} \phi_i(z_i)$.
Show for finite $\mathcal{Z}$: out of all probability distribution, $p(z)$, that fulfill $\mathbb{E}_{z\sim p(z)}[\phi(z)] = \mu_i$ for $i = 1, \dots, D$, the one with *highest entropy* has the form

$$p(z) \propto \exp\Big(\sum_i \theta_i \phi_i(z)\Big) \qquad \text{for some values } \theta_1, \dots, \theta_D \in \mathbb{R}.$$